A fin di bene: il nuovo potere della ragione artificiale

Stefano Isola

Un nuovo regno del bene artificiale?

Obsolescenza umana: dal controllo al controllo del controllo

Non ci interessa fare prodotti, ci interessa sfondare i limiti dell'inimmaginabile.

Sergey Brin (Google-Alphabet)

Siamo incapaci di farci un'immagine di ciò che noi stessi siamo stati capaci di fare. In questo senso siamo "utopisti a rovescio": mentre gli utopisti non sanno produrre ciò che concepiscono, noi non sappiamo immaginare ciò che abbiamo prodotto.

Günther Anders

apologie…

Non servirà più lavorare. Una delle grandi sfide del futuro sarà trovare un significato alla vita.

Elon Musk

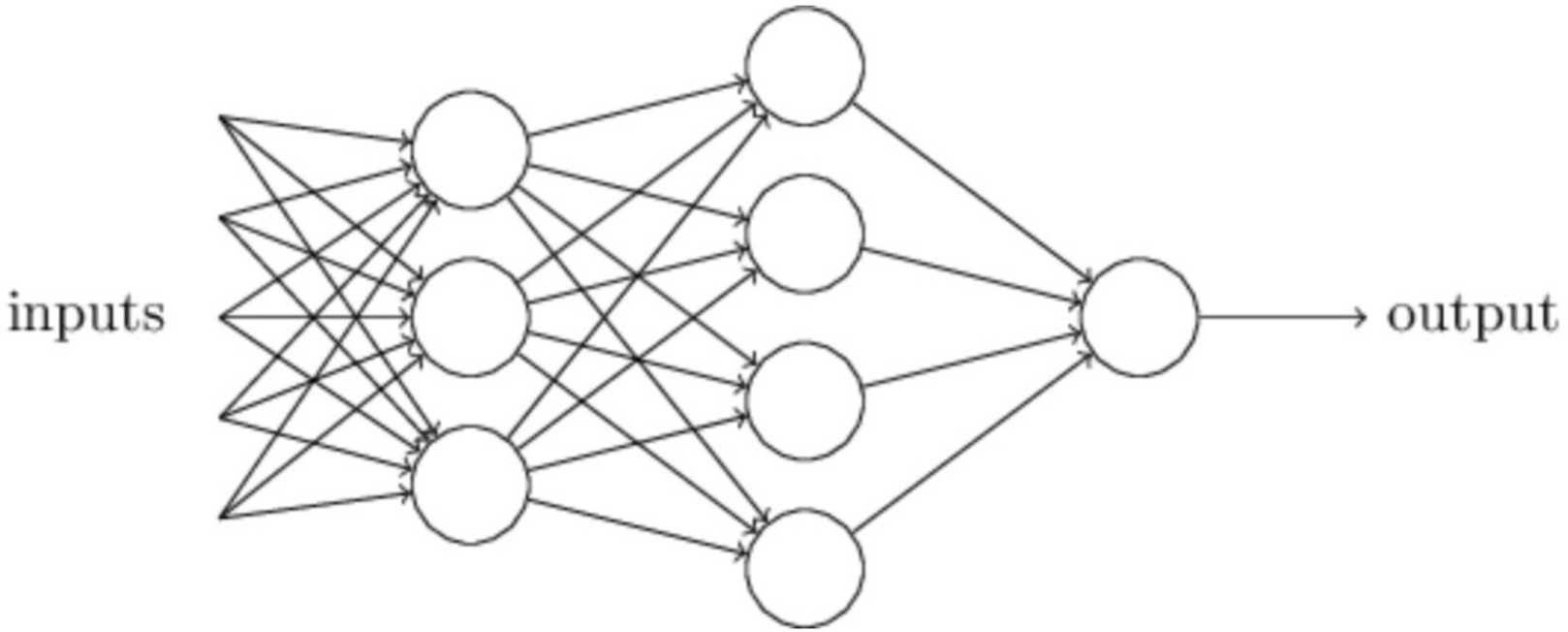
L'IA migliorerà il benessere umano e promuoverà la pace e la prosperità.

Dichiarazione finale del *AI safety summit* (Londra, 1-2 nov 2023)

La grande quantità di dati a disposizione rende il metodo scientifico obsoleto… i petabyte ci consentono di dire "la correlazione è sufficiente", possiamo smettere di cercare modelli.

Chris Anderson

Cognizione umana: dai tentativi di modellizzazione simbolica all'apprendimento automatico



Percettrone a due livelli

Correlazioni spurie nei *big data*

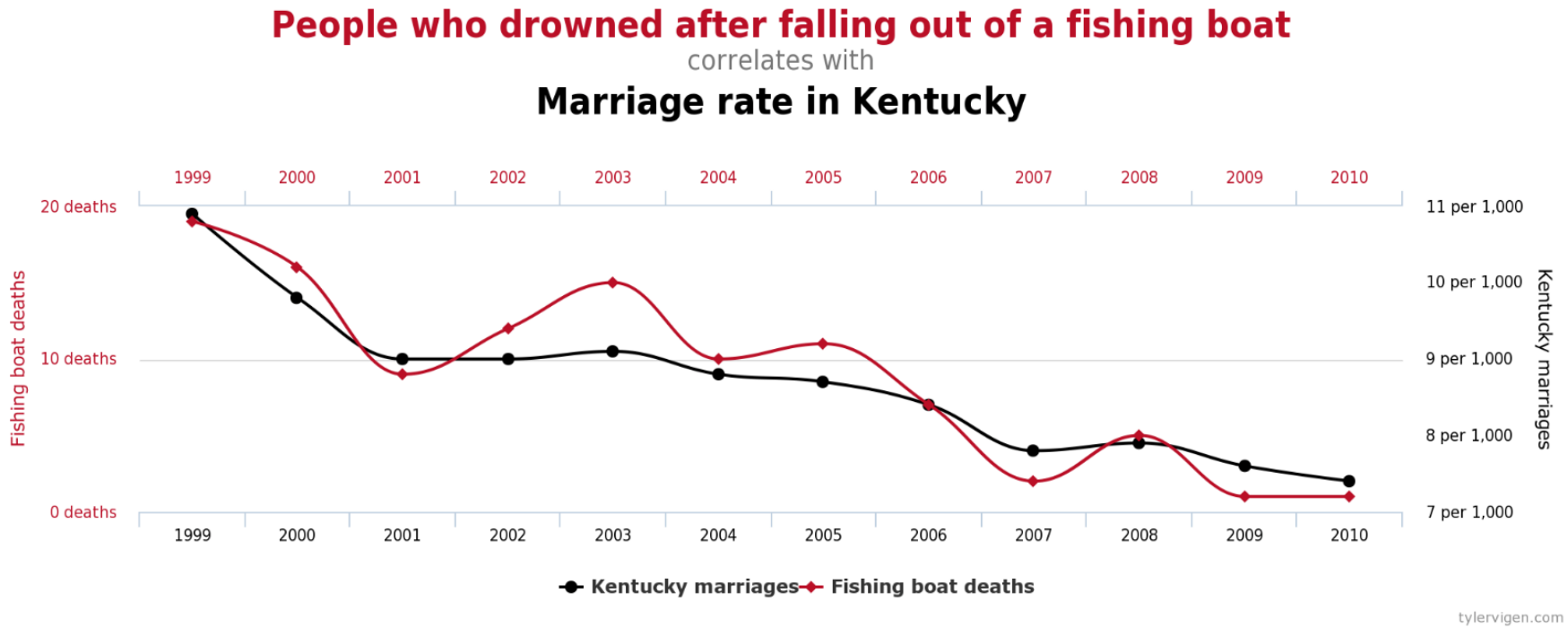


Figure 1: Spurious correlation with $r = 0.952407$.

… e allarmi

I sistemi di intelligenza artificiale in competizione con l'intelligenza umana possono comportare gravi rischi per la società e l'umanità.

Elon Musk *et al*

Se ci fosse un piano per la sopravvivenza della Terra, se solo approvassimo una moratoria di sei mesi, lo sosterrei. […] La moratoria sui nuovi grandi percorsi di addestramento deve essere indefinita e mondiale. Non ci possono essere eccezioni, nemmeno per i governi o le forze armate. Se la procedura iniziasse dagli Stati Uniti, la Cina dovrebbe capire che gli Stati Uniti non stanno cercando un vantaggio, ma piuttosto di impedire una tecnologia terribilmente pericolosa che non può avere un vero proprietario e che ucciderà tutti negli Stati Uniti, in Cina e sulla Terra.

Eliezer Yudkowsky

Scenari: ecocidio industriale e sua redenzione

Oltre a minacce ben note come l'olocausto nucleare, le prospettive di tecnologie radicalmente trasformatrici come i sistemi nanotecnologici e l'intelligenza artificiale ci presentano ***opportunità*** *e* ***rischi*** *senza precedenti.*

Non dovremmo incolpare la civiltà o la tecnologia per aver imposto grandi rischi esistenziali. Senza la tecnologia, le nostre possibilità di evitare rischi esistenziali sarebbero pari a zero. Con la tecnologia abbiamo qualche possibilità, anche se oggi i rischi maggiori risultano essere quelli generati dalla tecnologia stessa.

Nick Bostrom

Scenari: guerra robotica e auto-colonizzazione

Non possiamo permetterci un livellamento del vantaggio tecnologico. È indispensabile che il Dipartimento incoraggi la ricerca sulle tecnologie emergenti per evitare sorprese tecnologiche. Dobbiamo approfittare delle tecnologie commerciali all'avanguardia il cui rapido progresso è in grado di migliorare le nostre capacità militari.

Heidi Shyu (sottosegretario della Difesa USA con delega alla ricerca e l'ingegneria)

Produrre armi come i vaccini! Ursula von der Leyen

L'obiettivo è una guerra senza fine, non una guerra di successo.

Julian Assange (sul significato della guerra in Afghanistan)

Scenari: governo algoritmico e regressione civile

Le briciole di pane digitali che ci lasciamo dietro durante la nostra vita quotidiana – che rivelano più cose su di noi di qualsiasi cosa decidiamo di rivelare – forniscono un potente strumento per affrontare i problemi sociali.

Alex Pentland

La sorveglianza sociale è un cliente di prima qualità per l'intelligenza artificiale.

Eric Schmidt

Pensiero, linguaggio, memoria

Un cervello, o una mente, possono contenere rappresentazioni simboliche solo in virtù del rapporto tra il loro possessore e una comunità di persone che usa quei simboli [...] ne consegue che solo un essere inserito in una comunità linguistica può avere coscienza di sé.

John Dupré

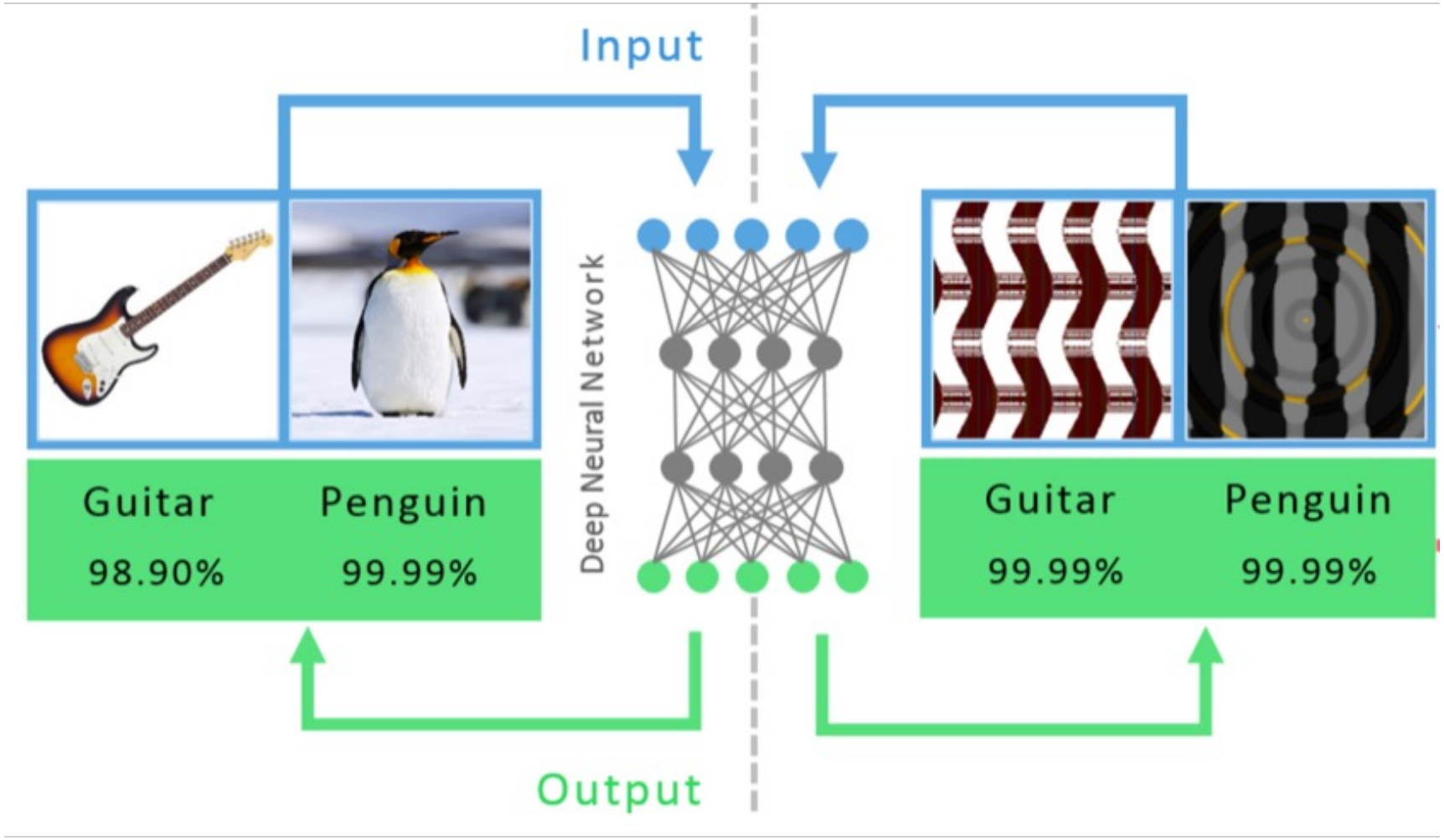
Molte cose devono già essere pronte nel linguaggio perché il puro denominare abbia un senso.

Ludwig Wittgenstein

Tra gli esseri umani c'è una tendenza universale a concepire ogni cosa come se stessi, ed a trasferire su ogni oggetto quelle qualità che ci sono familiari e delle quali siamo intimamente coscienti.

David Hume

Allucinazioni algoritmiche e disallineamento




TVBS新聞
TVBS

Principio di stabilità per l'IA e il correlato adattamento all'IA del mondo

Gli algoritmi complessi funzionano bene in situazioni le cui variabili siano ben definite, controllabili, e soprattutto stabili nel tempo. Solo in queste circostanze l'elaborazione di grandi quantità di dati, necessariamente riferiti al passato, può essere efficace per riconoscere o predire cose ed eventi nel futuro. Dunque, la strada per rendere efficaci prestazioni al momento inaffidabili dell'IA a causa dell'instabilità dell'ambiente (visione, guida automatica, screening di massa, etc.) è quella di indurre forme diffuse di adattamento: rendere l'ambiente in cui opera sempre più stabile, e tutti i processi che vi avvengono, tra cui il comportamento umano, sempre più prevedibili.